# Equation by equation estimation of the semi-diagonal BEKK model with covariates

Le Quyen Thieu

University Pierre and Marie Curie

1 September 2016

# Equation by equation estimation of the semi-diagonal BEKK model with covariates

Le Quyen Thieu,*

Université Pierre et Marie Curie.

**Abstract**

This paper provide the asymptotic normality of the Equation by Equation estimator for the semi-diagonal BEKK models augmented by the exogenous variables. The results are obtained without assuming that the innovations are independent, which allows investigate different additional explanatory variables into the information set.

*Keywords:* BEKK-X, Equation by equation estimation, exogenous variables, covariates, semi-diagonal BEKK-X

## 1 Introduction

Volatility modeling plays an crucial role in the study of financial mathematics, economics and statistics. Understanding volatilities of financial asset returns is important in hedging, risk management, and portfolio optimization. The family of GARCH has been widely used to model and forecast volatility (see Bollerslev and Wooldridge (1992) for a comprehensive review). In particular, multivariate GARCH models are popular for taking into account financial volatilities co-movements by estimating a conditional covariance

*Corresponding author: Le Quyen Thieu, Université Pierre et Marie Curie, France. Telephone: (+33)7 81 95 25 89. E-mail: thieulequyen1411@gmail.com

matrix. Recent developments in the estimation of multivariate second moment specifications include, among others, the constant conditional correlation (CCC) model of Bollerslev (1990), the BEKK model of Engle and Kroner (1995), the Factor GARCH model of Engle and Rothschild (1990), and the Dynamic Conditional Correlation (DCC) model of Engle (2002). One of the challenges in modeling second moments is to ensure the positive definiteness of the resultant covariance matrix without spuriously imposing a temporal pattern in conditional covariances. To cope with this issue, the BEKK model is one of the most flexible GARCH models that guarantees the time-varying covariance matrix to be positive definite by construction. However, the conditional covariance matrix is only explained by the past returns and volatilities while, in practice, many extra information, under the form of exogenous variables, can help explaining and forecasting financial volatility. For example, stock or portfolio covariances may be dependent on own financial characteristics such as liquidity measures, profitability measures, or valuation measures like price-to-book or cash-flow-to-book. In addition, overall market forces such as general credit market conditions might impact conditional covariances. It is, therefore, natural to ask for possible extensions of time series models to accommodate the wealth of information. Engle and Kroner (1995) suggest the BEKK model augmented by exogenous covariates (so-called BEKK-X). However, they only provide the estimation the BEKK model without impact of the covariates. Thieu (2016) presents the variance targeting estimation (VTE) of the BEKK-X model and establishes the Consistency and Asymptotic Normality (CAN) of the VTE. The BEKK-X model is also related to recent literature on GARCH models extended by additional explanatory variables (GARCH-X) with the aim of explaining and forecasting the volatility. Examples of such GARCH-X models include, amongst other, the heavy model of Shephard and Sheppard (2010), the GARCH-X$(1,1)$ of Han and Kristensen (2014), the Power ARCH$(p,q)$-X model of Francq and Thieu (2015), the Multivariate Log-GARCH-X model of Francq and Sucarrat (2015).

Another challenge in modeling a covariance matrix is the "curse of dimensionality", in particular, in the presence of exogenous variables. Indeed, when the dimension of the time series is large, or when the number of covariates is large, the number of parameters can become very large in MGARCH models. The log-likelihood function may therefore contain a numerous number of local maxima, and different starting-values may thus lead to different outcomes. A way of reducing this dimensionality curse is via so-called targeting which imposes a structure on the model intercept based on sample information. Variance targeting estimation is originally proposed by Engle and Mezrich (1996). Pedersen and Rahbek (2014) and Thieu (2015) consider this estimation method for the BEKK model and the BEKK-X model, respectively. Despite the potential benefits of the VT estimation, it remains some difficulties. Enforcing positive definiteness of the conditional covariance matrices the model with targeting implies a set of model constraints that are nonlinear in parameters. This becomes very complicated, except for the scalar case. Moreover, an important question is how much we lose in terms of statistical fit in moving from a BEKK/BEKK-X model to a restricted model with fewer parameters to estimate. Finally, in presence of covariates, the curse of dimensionality is still problematic.

In the present paper, I consider an approach that is so-called equation by equation (EbE) estimation to the semi-diagonal BEKK model with presence of the exogenous variables. This statistical method is initially proposed by Engle and Sheppard (2001) and Engle (2002) in the context of DCC models. The EbE method deals very effectively with high-dimensional problems and computationally costly log-likelihood. The asymptotic results for the EbE estimator (EbEE) of the volatility parameters, based on Quasi-Maximum Likelihood (QML) are developed in Francq and Zakoïan (2016). In their frame work, they only provide the asymptotic normality of the volatility parameter of the semi-diagonal without covariates. The asymptotic distribution of the estimator of the matrix intercept has not been given. The first goal of the present paper is to establish the strong consistency and the asymptotic distribution for the EbEE of the individual volatilities parameters of the semi-diagonal BEKK-X model. The second goal is to provide asymptotic results for the estimators of all matrix parameters.

The rest of paper is organized as follows. Section 2 introduces the BEKK model augemented with exogenous variables and presents the EbE method. The consistency and asymptotic behavior of the EbE estimators are investigated in Section 3. Section 4 contains the numerical illustrations. Section 5 concludes. All auxiliary lemmas and mathematical proofs are contained in Section 6.

Some notation and definition throughout the paper: For $m, n \in \mathbb{N}$, $I_m$ denotes the $(n \times n)$ identity matrix and $O_{m\times n}$ denotes the $(m \times n)$ zero matrix. The trace of $A$ is denoted by $tr(A)$, the determinant is denoted by $det(A)$ and the spectral radius of $A$ is denote by $\rho(A)$, *i.e.,* $\rho(A)$ is the maximum among the absolute values of the eigenvalues of $A$. The operator *vec* stacks all columns of a matrix into a column vector, *vech* denotes the operator that stacks only the lower triangular part including the diagonal of a symmetric matrix into a vector, and $vech^0$ is the operator which stacks the sub-diagonal elements (excluding the diagonal) of a matrix. The $(mn \times mn)$ commutation matrix $M_{mn}$ is defined such that, for any $(m \times n)$ matrix $A$, $M_{mn}vec(A) = vec(A')$. $D_m$ and $L_m$ denote the duplication matrix and elimination matrix defined such that , for any symmetric $(m \times m)$ matrix $A$, $vec(A) = D_m vech(A)$ and $vech(A) = L_m vec(A)$. Denote $T_m$ be a $m \times m(m+1)/2$ matrix such that $T_m vec(A) = diag(A)$ for any $m \times m$ matrix $A$. Let also $P_m$ be a $(\frac{m(m-1)}{2} \times m^2)$ matrix such that $vech^0(A) = P_m vec(A)$, for any symmetric $(m \times m)$ matrix $A$. The Kronecker product of $A$ and $B$ is defined by $A \otimes B = \{a_{ij}B\}$. The Euclidean norm of the matrix, or vector $A$, is defined as $\|A\| = \sqrt{tr(A'A)}$, and the spectral norm is defined as $\|A\|_{sp} = \sqrt{\rho(A'A)}$.

# 2 The model and EbE estimation

## 2.1 The model

Let $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \cdots, \varepsilon_{mt})'$ denote a $(m \times 1)$ vector of random variables and let $\boldsymbol{x}_t = (x_{1t}, \cdots, x_{rt})'$ be a $r$-dimentional vector of exogenous variables. Assume the existence of the $(m \times m)$ positive definite matrix $\boldsymbol{H}_t$ such that

$$E(\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1}) = \mathbf{0}, \qquad E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}) = \boldsymbol{H}_t, \tag{1}$$

where $\mathcal{F}_t = \sigma\{\boldsymbol{\varepsilon}_u, \boldsymbol{x}_{u'}'; u, u' \leq t\}$ implies the information set at time $t$. Note that $\boldsymbol{H}_t$ is the conditional covariance of $\boldsymbol{\varepsilon}_t$ given $\mathcal{F}_{t-1}$ the information set until time $t-1$.

We consider the following model

$$\begin{cases} \boldsymbol{\varepsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t \\ \boldsymbol{H}_t = \boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}'\boldsymbol{A}' + \boldsymbol{B}\boldsymbol{H}_{t-1}\boldsymbol{B} + \boldsymbol{C}\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\boldsymbol{C}' \end{cases} \tag{2}$$

where $\boldsymbol{A} = (a_{k\ell})_{1\leq k,\ell\leq m}$, $\boldsymbol{B} = diag(b_1, \cdots, b_m)$, $\boldsymbol{C} = (c_{k\ell})_{1\leq k\leq m, 1\leq \ell\leq r}$ and $\boldsymbol{\Omega} = (\omega_{k\ell})_{1\leq k,\ell\leq m}$ is a positive definite $(m \times m)$ matrix.

When there is no covariate (2) is called the semi-diagonal BEKK (see Francq and Zakoïan (2016)). Our model (2) can be so-called the semi-diagonal BEKK-X model.

Throughout of the paper, the following assumptions are made

**A1:** $E(\boldsymbol{\eta}_t|\mathcal{F}_{t-1}) = \mathbf{0}, Var(\boldsymbol{\eta}_t|\mathcal{F}_{t-1}) = I_m$.

**A2:** $(\boldsymbol{\varepsilon}_t, \boldsymbol{x}_t)$ is strictly stationary and ergodic process.

**A3:** $E(\|\boldsymbol{\varepsilon}_t\|^2) < \infty$ and $E(\|\boldsymbol{x}_t\|^2) < \infty$.

**Remark 1** *In many multivariate GARCH models, it is usual to assume that $(\boldsymbol{\eta}_t)$ is an iid white noise vector with zero mean and identity variance matrix. Comte and Lieberman (2003) and Pedersen and Rahbek (2014) establish the CAN of QMLE and VTE, respectively, for the BEKK model under this assumption. Francq and Zakoïan (2016) also provide the CAN of the EbEE for the semi-diagonal BEKK without covariate also under the assumption that the innovation process is iid. However, in presence of the exogenous variables, Assumption* **A1** *is weaker and seems more flexible than the iid assumption as several information sets $\mathcal{F}_t$ with different explanatory variables can be investigated. In univariate case, Francq and Thieu (2015) study the APARCH-X model under an assumption like* **A1**.

Denoting by $\sigma^2_{kt}$ the $k$-th diagonal element of $\boldsymbol{H}_t$, that is the variance of the $k$-th component, $\varepsilon_{kt}$, of $\boldsymbol{\varepsilon}_t$ conditional on $\mathcal{F}_{t-1}$

$$Var(\varepsilon_{kt}|\mathcal{F}_{t-1}) = \sigma^2_{kt}.$$

Define $\boldsymbol{D}_t$ as the diagonal matrix containing the conditional variances $\sigma^2_{kt}$ , *i.e.* $\boldsymbol{D}_t = diag(\sigma_{1t}, \ldots, \sigma_{mt})$ and let $\boldsymbol{\eta}^*_t = \boldsymbol{D}_t^{-1}\boldsymbol{\varepsilon}_t$ be the standardized returns. From (1), we have $E(\boldsymbol{\eta}^*_t|\mathcal{F}_{t-1}) = 0$ and $Var(\boldsymbol{\eta}^*_t|\mathcal{F}_{t-1}) = \boldsymbol{D}_t^{-1}\boldsymbol{H}_t\boldsymbol{D}_t^{-1}$. It follows that the components $\eta^*_{kt}$ of $\boldsymbol{\eta}^*_t$ satisfy, for $k = 1, \ldots, m$,

$$E(\eta^*_{kt}|\mathcal{F}_{t-1}) = 0, \qquad Var(\eta^*_{kt}|\mathcal{F}_{t-1}) = 1. \tag{3}$$

The individual volatilities are then parameterized as follows

$$\begin{cases} \varepsilon_{kt} = \sigma_{kt}\eta^*_{kt}, \\ \sigma^2_{kt} = \omega_{kk} + \left(\sum\limits_{\ell=1}^{m} a_{k\ell}\varepsilon_{\ell,t-1}\right)^2 + b_k^2\sigma^2_{k,t-1} + \left(\sum\limits_{s=1}^{r} c_{ks}x_{s,t-1}\right)^2. \end{cases} \tag{4}$$

To ensure the positivity of the volatilities, we assume that $\omega_{kk} > 0$. In view of (3), the process $(\boldsymbol{\eta}_t^*)$ can be called the vector of equation by equation (EbE) innovations of $(\boldsymbol{\varepsilon}_t)$. Let $\underline{\boldsymbol{a}}_k^0 = (a_{k1}^0, ..., a_{km}^0)'$ and $\underline{\boldsymbol{c}}_k^0 = (c_{k1}^0, \ldots, c_{kr}^0)'$ be, respectively, the $k$-th row vectors of the matrices $\boldsymbol{A}_0$ and $\boldsymbol{C}_0$. Then the vector of unknown parameters involved in the $k$-th equation (4) can be denoted by $\boldsymbol{\theta}_0^{(k)} = (\omega_{kk}^0, \underline{\boldsymbol{a}}_k^{0'}, (b_k^0)^2, \underline{\boldsymbol{c}}_k^{0'})' \in \mathbb{R}^d, d = m+r+2$. It is clear that an identifiability condition must be required such that $\sigma_{kt}^2$ is invariant to a change of sign of the vectors $\underline{\boldsymbol{a}}_k^0$, $\underline{\boldsymbol{c}}_k^0$ and $b_k^0$. Without lost of generality, we can assume that $a_{k1}^0 > 0$, $b_k^0 > 0$ and $c_{k1}^0 > 0$, for $k = 1, \ldots, m$.

Let $\boldsymbol{\theta}^{(k)} = (\omega_{kk}, \underline{\boldsymbol{a}}_k', b_k^2, \underline{\boldsymbol{c}}_k')' = (\omega_{kk}, a_{k1}, ..., a_{km}, b_k, c_{k1}, \ldots, c_{kr})'$ be a generic parameter vector of the parameter space $\boldsymbol{\Theta}^{(k)}$ which is an any compact subset of

$$(0, +\infty)^2 \times \mathbb{R}^{m-1} \times [0, 1) \times (0, +\infty) \times \mathbb{R}^{r-1}.$$

## 2.2 Equation-by-equation estimation of parameters

Let $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ be observations of a process satifying the semi-diagonal BEKK-X representation (2) and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be observations of a process of the explanatory variables. For all $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$, we recursively define $\widetilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)})$ for $t = 1, \ldots, n$ by

$$\widetilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)}) = \omega_{kk} + \left(\sum_{\ell=1}^{m} a_{k\ell}\varepsilon_{\ell,t-1}\right)^2 + b_k^2\widetilde{\sigma}_{k,t-1}^2(\boldsymbol{\theta}^{(k)}) + \left(\sum_{s=1}^{r} c_{ks}x_{s,t-1}\right)^2 \tag{5}$$

with the arbitrary initial values $\widetilde{\boldsymbol{\varepsilon}}_0, \widetilde{\sigma}_0$ and $\boldsymbol{x}_0$. Let

$$\widetilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}) = \frac{1}{n}\sum_{t=1}^{n} \tilde{\ell}_{kt}(\boldsymbol{\theta}^{(k)}), \quad \tilde{\ell}_{kt}(\boldsymbol{\theta}^{(k)}) = \log\widetilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)}) + \frac{\varepsilon_{kt}^2}{\widetilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)})}.$$

The EbE estimator, denoted by $\hat{\boldsymbol{\theta}}_n^{(k)}$, of the true parameter vector $\boldsymbol{\theta}_0^{(k)}$ is defined as a measurable solution of the following equation

$$\widehat{\boldsymbol{\theta}}_n^{(k)} = \underset{\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}}{\arg\min}\, \widetilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}). \tag{6}$$

Let $\boldsymbol{\theta}_0 = \left(\boldsymbol{\theta}_0^{(1)'}, \ldots, \boldsymbol{\theta}_0^{(m)'}\right)'$. Note that $\boldsymbol{\theta}_0$ includes the diagonal elements of $\boldsymbol{\Omega}_0$ and all components of the matrices $\boldsymbol{A}_0, \boldsymbol{B}_0$ and $\boldsymbol{C}_0$. This parameter vector belongs to the parameter space $\boldsymbol{\Theta}^m = \boldsymbol{\Theta}^{(1)} \times \cdots \times \boldsymbol{\Theta}^{(m)}$, whose generic element is denoted by $\boldsymbol{\theta} = \left(\boldsymbol{\theta}^{(1)'}, \ldots, \boldsymbol{\theta}^{(m)'}\right)'$. The estimator of $\boldsymbol{\theta}_0$ is given by $\widehat{\boldsymbol{\theta}}_n = \left(\widehat{\boldsymbol{\theta}}_n^{(1)'}, \ldots, \widehat{\boldsymbol{\theta}}_n^{(m)'}\right)'$ which is the collection of the equation by equation estimators.

Once the EbEE estimators $\widehat{\boldsymbol{A}}_n$, $\widehat{\boldsymbol{B}}_n$ and $\widehat{\boldsymbol{C}}_n$ of the matrices $\boldsymbol{A}_0, \boldsymbol{B}_0$ and $\boldsymbol{C}_0$, respectively, are obtained, the matrix $\boldsymbol{\Omega}_0$ can be fully estimated as follows

$$vech^0(\widehat{\boldsymbol{\Omega}}_n) = vech^0\left(\widehat{\Sigma}_{\boldsymbol{\varepsilon}n} - \widehat{\boldsymbol{A}}_n\widehat{\Sigma}_{\boldsymbol{\varepsilon}n}\widehat{\boldsymbol{A}}'_n - \widehat{\boldsymbol{B}}_n\widehat{\Sigma}_{\boldsymbol{\varepsilon}n}\widehat{\boldsymbol{B}}_n - \widehat{\boldsymbol{C}}_n\widehat{\Sigma}_{\boldsymbol{x}n}\widehat{\boldsymbol{C}}'_n\right), \tag{7}$$

where $\widehat{\Sigma}_{\boldsymbol{\varepsilon}n} = \dfrac{1}{n}\sum_{t=1}^n \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t$ and $\widehat{\Sigma}_{\boldsymbol{x}n} = \dfrac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t\boldsymbol{x}'_t$ are the empirical estimators of the second order moment matrices $\Sigma_{\boldsymbol{\varepsilon}} = E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t)$ and $\Sigma_{\boldsymbol{x}} = E(\boldsymbol{x}_t\boldsymbol{x}'_t)$, respectively. The estimation of the model (2) is thus nothing else than the estimation of

$$\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}'_0, \boldsymbol{\gamma}'_{\boldsymbol{\varepsilon}0}, \boldsymbol{\gamma}'_{\boldsymbol{x}0})', \quad \boldsymbol{\gamma}_{\boldsymbol{\varepsilon}0} = vech(\Sigma_{\boldsymbol{\varepsilon}}), \quad \boldsymbol{\gamma}_{\boldsymbol{x}0} = vech(\Sigma_{\boldsymbol{x}}).$$

Its estimator can be given by $\widehat{\boldsymbol{\vartheta}}_n = (\widehat{\boldsymbol{\theta}}'_n, \widehat{\boldsymbol{\gamma}}'_{\boldsymbol{\varepsilon}n}, \widehat{\boldsymbol{\gamma}}'_{\boldsymbol{x}n})'$, where $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}n} = vech(\widehat{\Sigma}_{\boldsymbol{\varepsilon}n})$ and $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{x}n} = vech(\widehat{\Sigma}_{\boldsymbol{x}n})$.

## 3 EbE estimation inference

For the consistency of the estimator, the assumptions following will be made

**A4:** $\boldsymbol{\theta}_0^{(k)} \in \boldsymbol{\Theta}^{(k)}, \boldsymbol{\Theta}^{(k)}$ is compact, for $k = 1, \dots, m$.

**A5:** $\rho(\boldsymbol{A}_0 \otimes \boldsymbol{A}_0 + \boldsymbol{B}_0 \otimes \boldsymbol{B}_0) < 1$ and $\sum_{k=1}^m b_k^2 < 1$, for all $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$.

**A6:** There exists $s > 0$ such that $E|\varepsilon_{kt}|^s < \infty$ and $E|x_{kt}|^s < \infty$.

**A7:** For all $\ell^* = 1, \dots, m$, $\varepsilon^2_{\ell^* t}$ does not belong to the Hilbert space generated by the linear combinations of the $\varepsilon_{\ell u}\varepsilon_{\ell' u}$'s, the $x_{sv}x_{s'v}$'s for $u < t$, $v \leq t$, $\ell, \ell' = 1, \dots, m$, $s, s' = 1, \dots, r$ and the $\varepsilon_{\ell t}\varepsilon_{\ell' t}$ for $(\ell, \ell') \neq (\ell^*, \ell^*)$.

**A8:** For all $s^* = 1, \dots, r$, $x^2_{s^* t}$ does not belong to the Hilbert space generated by the linear combinations of the the $x_{sv}x_{s'v}$'s for $v < t$, $s, s' = 1, \dots, r$ and the $x_{st}x_{s't}$ for $(s, s') \neq (s^*, s^*)$.

**Remark 2** *Assumptions* **A7** *and* **A8** *are identification conditions. For simplicity, let us consider* (2) *when $m = 2$, $r = 2$ and the conditional covariance matrix is given by*

$$\boldsymbol{H}_t = \boldsymbol{\Omega} + \boldsymbol{A}\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}\boldsymbol{A}' + \boldsymbol{C}\boldsymbol{x}_{t-1}\boldsymbol{x}'_{t-1}\boldsymbol{C}', \tag{8}$$

*where* $\mathbf{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}$ *is a symmetric positive definite matrix,* $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ *and* $\boldsymbol{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$*, with* $a_{11} > 0$*,* $a_{21} > 0$*,* $c_{11} > 0$ *and* $c_{21} > 0$*. The volatility of the first component,* $\varepsilon_{1t}$*, of* $\boldsymbol{\varepsilon}_t$ *is thus given by*

$$\sigma_{1t}^2 = \omega_{11}+a_{11}\varepsilon_{1,t-1}^2+2a_{11}a_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1}+a_{12}\varepsilon_{2,t-1}^2+c_{11}x_{1,t-1}^2+2c_{11}c_{12}x_{1,t-1}x_{2,t-1}+c_{12}x_{2,t-1}^2.$$

*Assumption* **A7** *precludes, for example, that* $x_{1,t-1} = \varepsilon_{1,t-1}$ *for which the model is not identifiable.*

*Similarly, Assumption* **A8** *rules out the existence of linear combination of a finite number of the* $x_{s,t-i}x_{s',t-i}$ *that is obviously necessary for the identifiability.*

**Theorem 1** *Under* **A1 - A8***, the EbEE of* $\boldsymbol{\theta}_0^{(k)}$ *is strongly consistent*

$$\widehat{\boldsymbol{\theta}}_n^{(k)} \to \boldsymbol{\theta}_0^{(k)}, \text{ a.s. as } n \to \infty.$$

The following result is an immediate consequence of Theorem 1.

**Corollary 1** *Under the assumptions of Theorem 1,* $\widehat{\boldsymbol{\vartheta}}_n$ *is a strongly consistent estimator of* $\boldsymbol{\vartheta}_0$*.*

Now we turn to the asymptotic distribution of the estimation. We need some additional assumptions.

**A9:** $\boldsymbol{\theta}_0^{(k)}$ belongs to the interior of the parameter space $\mathbf{\Theta}^{(k)}$, for $k = 1, \ldots, m$.

**A10:** $E\|\boldsymbol{\eta}_t\|^{4(1+\delta)} < \infty, E||\boldsymbol{\varepsilon}_t||^{4(1+1/\delta)} < \infty$ and $E||\boldsymbol{x}_t||^{4(1+1/\delta)} < \infty$ for some $\delta > 0$.

**A11:** The process $\boldsymbol{z}_t = (\boldsymbol{x}_t', \boldsymbol{\varepsilon}_t', \boldsymbol{\eta}_t')'$ satisfies $E\|\boldsymbol{\varepsilon}_t\|^{(4+2\nu)(1+1/\delta)} < \infty$, $E\|\boldsymbol{x}_t\|^{(4+2\nu)(1+1/\delta)} < \infty$ and $E\|\boldsymbol{\eta}_t\|^{(4+2\nu)(1+\delta)} < \infty$, moreover the strong mixing coefficients, $\alpha_{\boldsymbol{z}}(h)$, of the process $(\boldsymbol{z}_t)$ are such that

$$\sum_{h=0}^{\infty} \{\alpha_{\boldsymbol{z}}(h)\}^{\nu/(2+\nu)} < \infty \text{ for some } \nu > 0 \text{ and } \delta > 0.$$

Let $\underline{\boldsymbol{H}}_{t,s}(\boldsymbol{\vartheta})$ be such that, for $s > 0$,

$$vec(\underline{\boldsymbol{H}}_{t,s}(\boldsymbol{\vartheta})) = \sum_{k=0}^{s}(\boldsymbol{B}^{\otimes 2})^k \left(vec(\boldsymbol{\Omega}) + \boldsymbol{A}^{\otimes 2}vec(\boldsymbol{\varepsilon}_{t-k-1}\boldsymbol{\varepsilon}'_{t-k-1}) + \boldsymbol{C}^{\otimes 2}vec(\boldsymbol{x}_{t-k-1}\boldsymbol{x}'_{t-k-1})\right),$$

where $A^{\otimes 2}$ denotes the Kronecker product of a matrix $A$ and itself. Let also $\mathcal{S}$ be a subspace such that for all $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$, $\boldsymbol{H}_t(\boldsymbol{\vartheta}) \in \mathcal{S}$ and for all $s > 0$, $\underline{\boldsymbol{H}}_{t,s}(\boldsymbol{\vartheta}) \in \mathcal{S}$.

**A12:** There exists $K > 0$ such that

$$\left\|\boldsymbol{H}_t^{1/2}(\boldsymbol{\vartheta}) - \boldsymbol{H}_t^{*1/2}(\boldsymbol{\vartheta})\right\| \le K \left\|\boldsymbol{H}_t(\boldsymbol{\vartheta}) - \boldsymbol{H}_t^*(\boldsymbol{\vartheta})\right\| \quad \text{for all } \boldsymbol{H}_t(\boldsymbol{\vartheta}), \boldsymbol{H}_t^*(\boldsymbol{\vartheta}) \in \mathcal{S}.$$

**Remark 3** *Assumptions* **A11** *and* **A12** *are also required in Thieu (2016) for the BEKK-X model estimated by the variance targeting method.*

In order to state the asymptotic normality, we have to introduce the following notations.

Let the $(d \times d)$ matrices $\boldsymbol{J}_{ks} = E(\Delta_{kt}\Delta'_{st})$, $k, s = 1, \dots m$, where $\Delta_{kt} = \dfrac{1}{\sigma_{kt}^2}\dfrac{\partial\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\partial\boldsymbol{\theta}^{(k)}}$ and let $\boldsymbol{J} = diag\{\boldsymbol{J}_{11}, \dots, \boldsymbol{J}_{mm}\}$ in bloc-matrix notation. Let also $\boldsymbol{\Delta}_t = diag(\Delta_{1t}, \dots, \Delta_{mt})$, $\boldsymbol{N}_1 = L_m(I_{m^2} - \boldsymbol{A}_0^{\otimes 2} - \boldsymbol{B}_0^{\otimes 2})^{-1}(I_{m^2} - \boldsymbol{B}_0^{\otimes 2})$ and $\boldsymbol{N}_2 = L_m(I_{m^2} - \boldsymbol{A}_0^{\otimes 2} - \boldsymbol{B}_0^{\otimes 2})^{-1}\boldsymbol{C}_0^{\otimes 2}D_r$. We also define the following matrices

$$\boldsymbol{\Sigma}_{11} = \sum_{h=-\infty}^{\infty} cov\left(\boldsymbol{\Upsilon}_{0t}vec\left(\boldsymbol{\eta}_t\boldsymbol{\eta}'_t\right), \boldsymbol{\Upsilon}_{0,t-h}vec\left(\boldsymbol{\eta}_{t-h}\boldsymbol{\eta}'_{t-h}\right)\right), \tag{9}$$

$$\boldsymbol{\Sigma}_{22} = \sum_{h=-\infty}^{\infty} cov\left(vech(\boldsymbol{x}_t\boldsymbol{x}'_t), vech(\boldsymbol{x}_{t-h}\boldsymbol{x}'_{t-h})\right), \tag{10}$$

$$\boldsymbol{\Sigma}_{12} = \sum_{h=-\infty}^{\infty} cov\left(vech(\boldsymbol{x}_t\boldsymbol{x}'_t), \boldsymbol{\Upsilon}_{0,t-h}vec\left(\boldsymbol{\eta}_{t-h}\boldsymbol{\eta}'_{t-h}\right)\right), \tag{11}$$

where

$$\boldsymbol{\Upsilon}_{0t} = \begin{pmatrix} \boldsymbol{\Delta}_t T_m\left(\boldsymbol{D}_{0t}^{-1}\boldsymbol{H}_{0t}^{1/2}\right) \otimes \left(\boldsymbol{D}_{0t}^{-1}\boldsymbol{H}_{0t}^{1/2}\right) \\ \boldsymbol{H}_{0t}^{1/2} \otimes \boldsymbol{H}_{0t}^{1/2} \end{pmatrix}. \tag{12}$$

**Theorem 2** *Under* **A1 - A12***, as $n \to \infty$,*

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\gamma}_{\boldsymbol{\varepsilon}n} - \gamma_{\boldsymbol{\varepsilon}0} \\ \widehat{\gamma}_{\boldsymbol{x}n} - \gamma_{\boldsymbol{x}0} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}'\right), \tag{13}$$

*where*

$$\boldsymbol{\Gamma} = \begin{pmatrix} -\boldsymbol{J}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{N}_1 & \boldsymbol{N}_2 \\ \mathbf{0} & \mathbf{0} & I_{r(r+1)/2} \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \tag{14}$$

The original parameter vector is denoted by

$$\boldsymbol{\xi}_0 = \left((vech^0(\boldsymbol{\Omega}_0))', \boldsymbol{\theta}_0'\right)' \in \mathbb{R}^{m(m-1)/2+md}. \tag{15}$$

The estimator of $\boldsymbol{\xi}_0$ can be given by $\widehat{\boldsymbol{\xi}}_n = \left(\widehat{\underline{\boldsymbol{\omega}}}_n', \widehat{\boldsymbol{\theta}}_n'\right)'$, where $\widehat{\underline{\boldsymbol{\omega}}}_n = vech^0(\widehat{\boldsymbol{\Omega}}_n)$ is the estimator of $\underline{\boldsymbol{\omega}}_0 = vech^0(\boldsymbol{\Omega}_0)$.

The strong consistency and the asymptotic distribution of the estimation is given as follows

**Theorem 3** *If* **A1 - A8** *hold, the estimator* $\widehat{\boldsymbol{\xi}}_n$ *of* $\boldsymbol{\xi}_0$ *is strongly consistent:*

$$\widehat{\boldsymbol{\xi}}_n \to \boldsymbol{\xi}_0 \text{ a.s. as } n \to \infty.$$

*If, in addition,* **A9 - A12** *hold, then*

$$\sqrt{n} \begin{pmatrix} \widehat{\underline{\boldsymbol{\omega}}}_n - \underline{\boldsymbol{\omega}}_0 \\ \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}'\right), \tag{16}$$

*where*

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_1 \\ \boldsymbol{\Omega}_2 \end{pmatrix}, \qquad \begin{aligned} \boldsymbol{\Omega}_1 &= \begin{pmatrix} \boldsymbol{A}^* & \boldsymbol{B}^* & \boldsymbol{C}^* & \boldsymbol{E}^* & \boldsymbol{X}^* \end{pmatrix} \boldsymbol{\Psi}, \\ \boldsymbol{\Omega}_2 &= \begin{pmatrix} I_{md} & \mathbf{0}_{m(m+1)/2} & \mathbf{0}_{r(r+1)/2} \end{pmatrix}, \end{aligned} \tag{17}$$

*with*

$$\begin{aligned}
\boldsymbol{A}^* &= -P_m \left\{ I_m \otimes (\boldsymbol{A}_0 \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) + ((\boldsymbol{A}_0 \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \otimes I_m) M_{mm} \right\}, \\
\boldsymbol{B}^* &= -P_m \left\{ I_m \otimes (\boldsymbol{B}_0 \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) + (\boldsymbol{B}_0 \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \otimes I_m \right\}, \\
\boldsymbol{C}^* &= -P_m \left\{ I_m \otimes (\boldsymbol{C}_0 \boldsymbol{\Sigma}_{\boldsymbol{x}}) + ((\boldsymbol{C}_0 \boldsymbol{\Sigma}_{\boldsymbol{x}}) \otimes I_m) M_{mr} \right\}, \\
\boldsymbol{E}^* &= P_m (I_{m^2} - \boldsymbol{B}_0 \otimes \boldsymbol{B}_0 - \boldsymbol{A}_0 \otimes \boldsymbol{A}_0) D_m, \quad \boldsymbol{X}^* = -P_m (\boldsymbol{C}_0 \otimes \boldsymbol{C}_0) D_r.
\end{aligned}$$

# 4 Numerical illustrations

This section presents the results of Monte Carlo simulations studies aimed at examining the performance of the EbEE of the semi-diagonal BEKK-X model and comparing the finite sample properties of the EbEE and VTE.

First, the quality of the EbEE will be evaluated through a Monte Carlo experiment. To keep the computation burden feasible, we focus on the bivariate case, $m = 2$, but the results should generalize in an obvious way to higher dimensions. The vector of the exogenous variables is $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}, \boldsymbol{x}_{2t})' = (\boldsymbol{z}_{t-1}, \boldsymbol{z}_{t-2})'$, where $\boldsymbol{z}_{t-1}$ and $\boldsymbol{z}_{t-2}$ are two lagged values of an APARCH(1,1)

$$\boldsymbol{z}_t = \sigma_t e_t, \qquad \sigma_t = 0.046 + 0.027\boldsymbol{z}_{t-1}^{+} + 0.092\boldsymbol{z}_{t-1}^{-} + 0.843\sigma_{t-1}, \tag{18}$$

where $\sqrt{2}e_t$ is i.i.d and follows a Student distribution with 4 degrees of freedom. As discussed in Remark 1, the $\boldsymbol{\eta}_t$'s do not need to be iid with zero mean and identity variance matrix. In this experiment, the $\boldsymbol{\eta}_t$'s are assumed to follow a bivariate Student distribution with $(4 + |\boldsymbol{x}_{1,t-1}|)$ degrees of freedom. The BEKK-X parameter are taken as follows

$$\boldsymbol{\Omega}_0 = \begin{pmatrix} 0.3 & 0.2 \\ 0.2 & 0.4 \end{pmatrix}, \boldsymbol{A}_0 = \begin{pmatrix} 0.15 & 0.1 \\ 0.1 & 0.2 \end{pmatrix}, \boldsymbol{B}_0 = \begin{pmatrix} 0.8 & 0.0 \\ 0.0 & 0.9 \end{pmatrix}, \boldsymbol{C}_0 = \begin{pmatrix} 0.15 & 0.05 \\ 0.1 & 0.2 \end{pmatrix}. \tag{19}$$

The data series are generated 500 times for $n = 1000$ and $n = 5000$ observations. For each data series, $(n + 500)$ observations of $\boldsymbol{\varepsilon}_t$ are simulated and then the first 500 observations are discarded in each simulation to minimize the effect of the initial values. The results of the simulation study are presented in Table 2. They are in accordance with the consistency of the EbEE, in particular the medians of the estimated parameters are close to the true values. As expected, the accuracy of the estimation increases as the sample size increases from $n = 1000$ to $n = 5000$.

Firgures 1, 2, 3 and 4 show non parametric estimators of the density of the components of $\widehat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0$. As expected, the estimated densities of the estimators over the 500 simulations are close to a Gaussian density for $n$ sufficiently large.

Next, a Monte Carlo experiment is performed with the aim to compare the empirical accuracies of the EbEE and VTE. The true parameter matrices are taken as same as in the previous experiment. $\boldsymbol{\eta}_t$ are assumed to be independent and normally distributed $\mathcal{N}(0, 1)$. Table 2 summarizes the distributions of the two estimators over 500 independent simulations of the model, for the length $n = 1000$, $n = 3000$ and $n = 5000$. From these results, the biases of the VTE are smaller than the one of the EbEE for $n = 1000$, $n = 3000$

Table 1: Sampling distribution of the EbEE of $\boldsymbol{\vartheta}_0$ over 500 replications for the BEKK-X(1,1) model

| parameter | true val. | bias | RMSE | min | $Q_1$ | $Q_2$ | $Q_3$ | max |
|---|---|---|---|---|---|---|---|---|
| | | | | $n = 1,000$ | | | | |
| $vec(\boldsymbol{\Omega})$ | 0.30 | 0.0100 | 0.1427 | 0.0228 | 0.2136 | 0.2880 | 0.3826 | 1.0031 |
| | 0.20 | 0.0872 | 0.2576 | -0.1338 | 0.1158 | 0.2442 | 0.3916 | 1.6704 |
| | 0.40 | 0.0328 | 0.3037 | 0.0000 | 0.2268 | 0.3775 | 0.5790 | 2.0654 |
| $\boldsymbol{A}$ | 0.15 | -0.0504 | 0.1244 | -0.3250 | 0.0269 | 0.1158 | 0.1910 | 0.3889 |
| | 0.10 | -0.0207 | 0.1871 | -0.6000 | -0.0121 | 0.1020 | 0.2039 | 0.6000 |
| | 0.10 | -0.0210 | 0.0864 | -0.2045 | 0.0552 | 0.0985 | 0.1294 | 0.2646 |
| | 0.20 | -0.0241 | 0.1197 | -0.4465 | 0.1289 | 0.1890 | 0.2535 | 0.4824 |
| $diag(\boldsymbol{B})$ | 0.80 | -0.0106 | 0.0617 | 0.5064 | 0.7577 | 0.7968 | 0.8287 | 0.9490 |
| | 0.90 | -0.0083 | 0.0371 | 0.6658 | 0.8735 | 0.8954 | 0.9153 | 0.9765 |
| $\boldsymbol{C}$ | 0.15 | -0.0023 | 0.0289 | -0.1420 | 0.1341 | 0.1494 | 0.1635 | 0.2115 |
| | 0.10 | 0.0081 | 0.0698 | -0.2306 | 0.0618 | 0.1080 | 0.1576 | 0.3110 |
| | 0.05 | -0.0008 | 0.0273 | -0.0432 | 0.0316 | 0.0482 | 0.0656 | 0.1575 |
| | 0.20 | -0.0162 | 0.0594 | -0.2287 | 0.1533 | 0.1927 | 0.2201 | 0.3573 |
| | | | | $n = 5,000$ | | | | |
| $vec(\boldsymbol{\Omega})$ | 0.30 | 0.0020 | 0.0530 | 0.1959 | 0.2631 | 0.2966 | 0.3308 | 0.5083 |
| | 0.20 | 0.0066 | 0.0829 | 0.0136 | 0.1479 | 0.2076 | 0.2622 | 0.4684 |
| | 0.40 | 0.0158 | 0.1171 | 0.1270 | 0.3354 | 0.4094 | 0.4790 | 0.8367 |
| $\boldsymbol{A}$ | 0.15 | -0.0007 | 0.0393 | 0.0004 | 0.1228 | 0.1515 | 0.1749 | 0.2602 |
| | 0.10 | 0.0049 | 0.0766 | 0.2135 | 0.0557 | 0.1036 | 0.1620 | 0.3341 |
| | 0.10 | -0.0007 | 0.0209 | 0.0390 | 0.0856 | 0.0995 | 0.1128 | 0.1635 |
| | 0.20 | -0.0044 | 0.0392 | 0.0828 | 0.1711 | 0.1950 | 0.2244 | 0.3111 |
| $diag(\boldsymbol{B})$ | 0.80 | -0.0025 | 0.0234 | 0.7033 | 0.7845 | 0.7990 | 0.8138 | 0.8536 |
| | 0.90 | -0.0023 | 0.0133 | 0.8466 | 0.8898 | 0.8986 | 0.9063 | 0.9353 |
| $\boldsymbol{C}$ | 0.15 | 0.0006 | 0.0099 | 0.1246 | 0.1444 | 0.1503 | 0.1569 | 0.1787 |
| | 0.10 | 0.0034 | 0.0310 | 0.0259 | 0.0820 | 0.1016 | 0.1217 | 0.2242 |
| | 0.05 | -0.0005 | 0.0110 | 0.0166 | 0.0418 | 0.0491 | 0.0571 | 0.0803 |
| | 0.20 | -0.0036 | 0.0234 | 0.0955 | 0.1844 | 0.1990 | 0.2113 | 0.2597 |

RMSE is the Root Mean Square Error, $Q_i$, $i = 1, 3$, denote the quartiles.

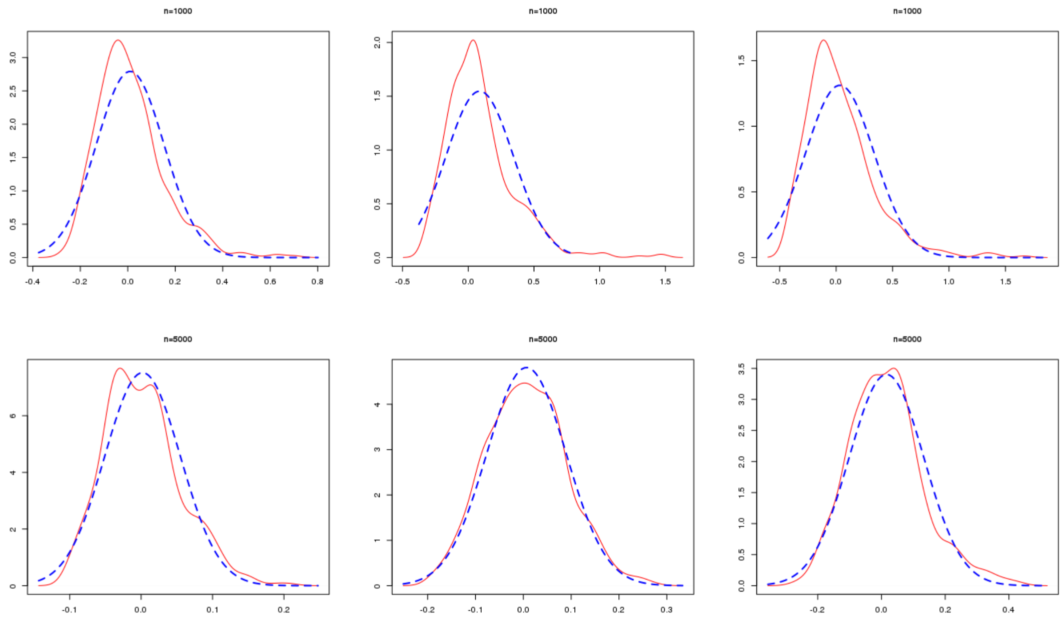

Figure 1: Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved $vech(\mathbf{\Omega})$.

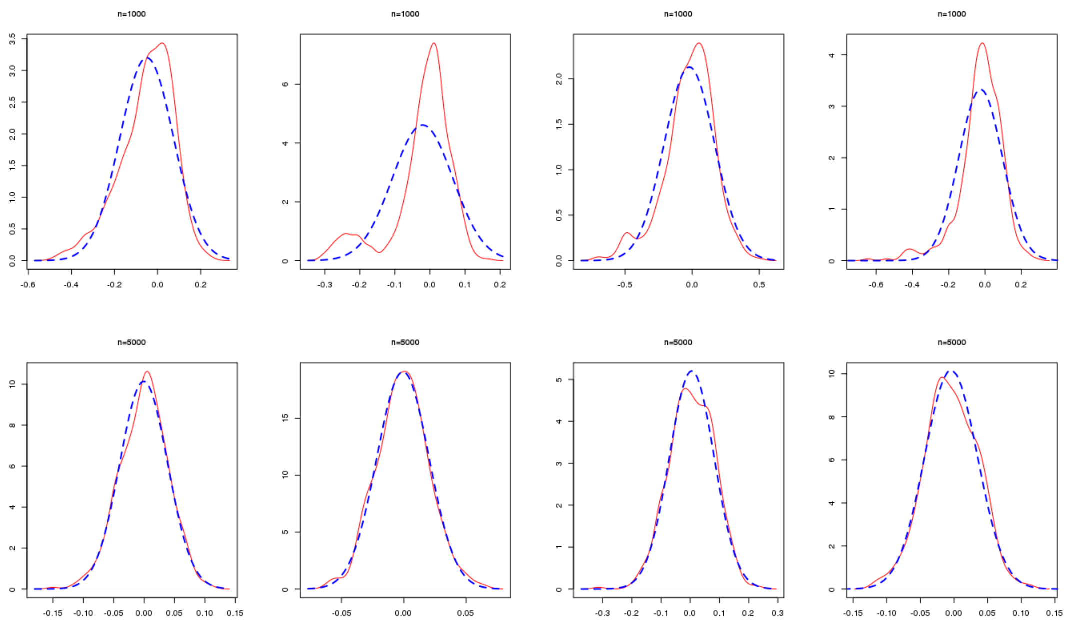

Figure 2: Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved $\boldsymbol{A}$.

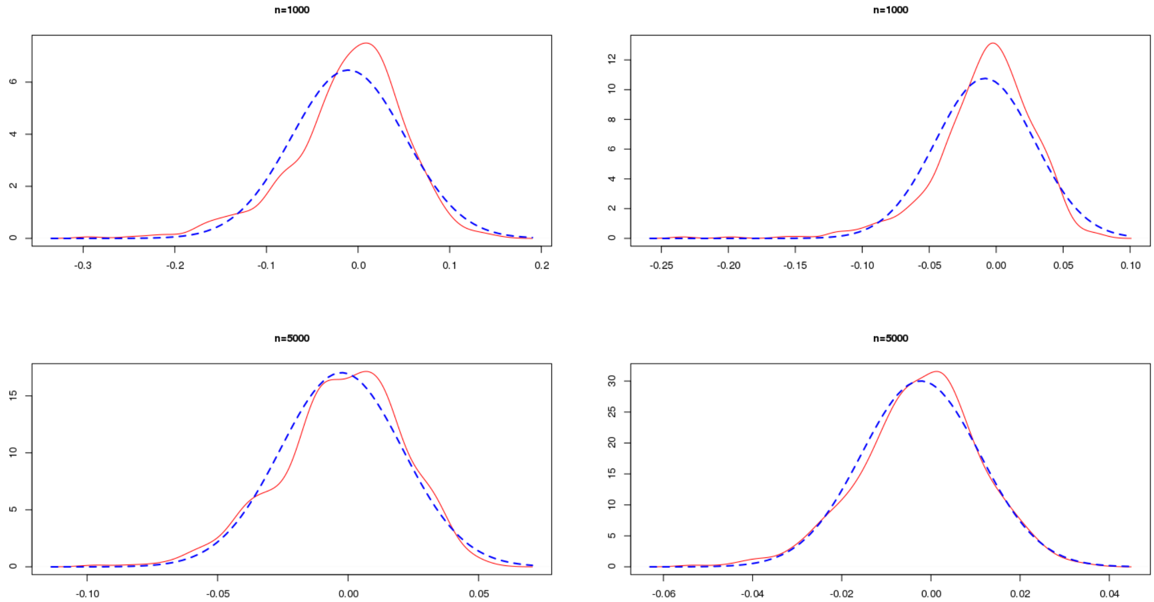


Figure 3: Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved $diag(\boldsymbol{B})$.

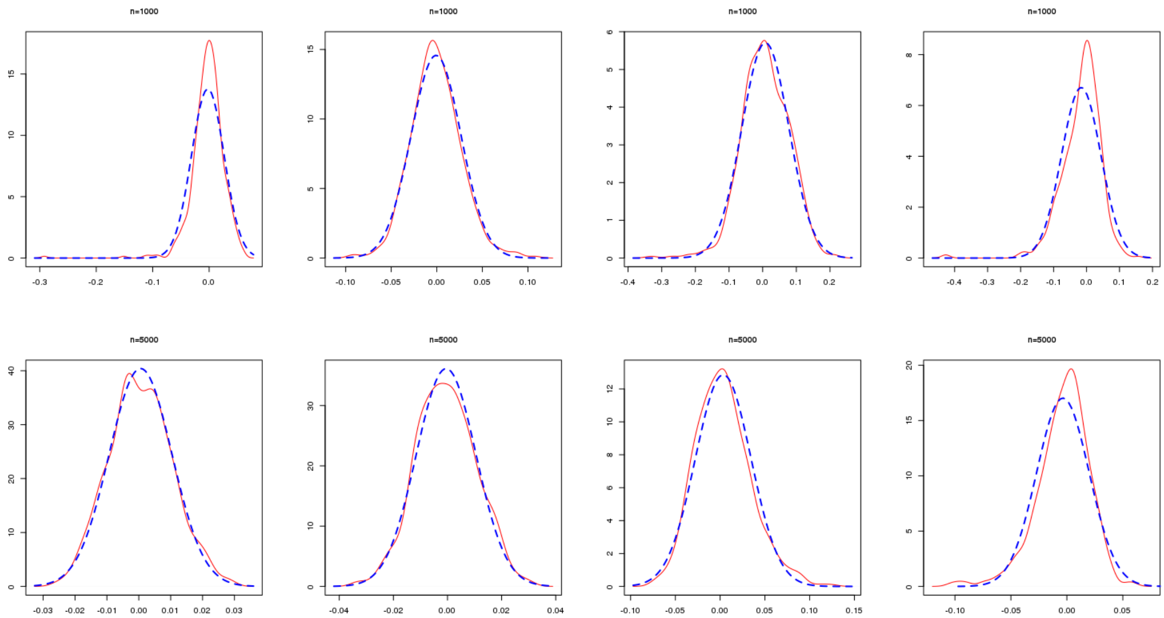


Figure 4: Kernel density estimator (in full line) of the distribution of the EbEE errors for the estimation of the parameters involved $\boldsymbol{C}$.

and $n = 5000$. But the differences are tiny when $n$ is sufficiently large, for example $n = 3000$ and $n = 5000$. Thus, the EbE approach does not seem to create a serious bias problem in the estimation of the dynamic parameters. Figure 5 displays the distribution of the estimation errors for simulations of length $n = 5000$. The upper-left, upper-right, bottom-left and bottom-right panels correspond respectively to the estimation errors for the parameters involved in $vech(\boldsymbol{\Omega})$, $\boldsymbol{A}$, $diag(\boldsymbol{B})$ and $\boldsymbol{C}$. The distributions of the EbEE and VTE are quite similar.

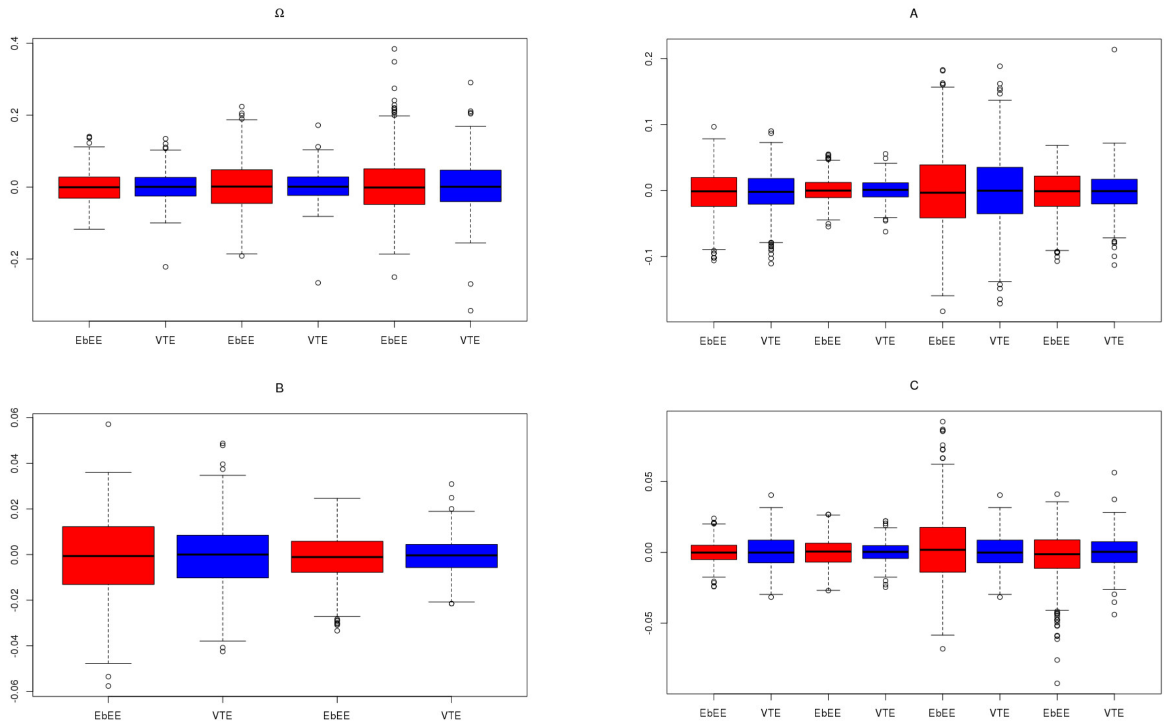


Figure 5: Boxplots of 500 estimation errors for the EbEE and VTE.

## 5 Conclusion

This paper suggest an effective approach, equation by equation estimation method, for estimation and inference on the semi-diagonal BEKK included the exogenous variables. This approach is recently widely used by researchers and practitioners, consisting in estimating separately the individual volatilities parameters in the first step, and estimating

Table 2: Sampling distribution of the EbEE and VTE of $\boldsymbol{\vartheta}_0$ over 500 replications for the BEKK-X(1,1) model

| | true | estim. | Median | bias | RMSE | Median | bias | RMSE | Median | bias | RMSE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $n = 1,000$ | | | $n = 3,000$ | | | $n = 5,000$ | | |
| $\boldsymbol{\Omega}$ | 0.30 | EbEE | 0.293 | 0.009 | 0.102 | 0.299 | 0.005 | 0.057 | 0.300 | 0.001 | 0.044 |
| | | VTE | 0.295 | -0.002 | 0.094 | 0.297 | 0.003 | 0.048 | 0.300 | 0.002 | 0.039 |
| | 0.20 | EbEE | 0.240 | 0.059 | 0.198 | 0.201 | 0.007 | 0.094 | 0.202 | 0.004 | 0.068 |
| | | VTE | 0.199 | 0.005 | 0.105 | 0.202 | 0.005 | 0.051 | 0.201 | 0.003 | 0.039 |
| | 0.40 | EbEE | 0.401 | 0.031 | 0.224 | 0.403 | 0.013 | 0.107 | 0.399 | 0.008 | 0.085 |
| | | VTE | 0.404 | 0.010 | 0.204 | 0.406 | 0.008 | 0.090 | 0.401 | 0.005 | 0.069 |
| $\boldsymbol{A}$ | 0.15 | EbEE | 0.140 | -0.025 | 0.110 | 0.151 | -0.001 | 0.047 | 0.149 | -0.003 | 0.034 |
| | | VTE | 0.142 | -0.010 | 0.079 | 0.150 | 0.000 | 0.042 | 0.148 | -0.003 | 0.032 |
| | 0.10 | EbEE | 0.094 | -0.012 | 0.168 | 0.108 | 0.008 | 0.083 | 0.097 | 0.000 | 0.059 |
| | | VTE | 0.109 | 0.012 | 0.140 | 0.109 | 0.008 | 0.074 | 0.100 | 0.001 | 0.054 |
| | 0.10 | EbEE | 0.097 | -0.014 | 0.066 | 0.098 | -0.003 | 0.029 | 0.100 | 0.001 | 0.018 |
| | | VTE | 0.100 | -0.001 | 0.040 | 0.098 | -0.002 | 0.024 | 0.101 | 0.001 | 0.017 |
| | 0.20 | EbEE | 0.194 | -0.030 | 0.108 | 0.195 | -0.008 | 0.050 | 0.199 | -0.003 | 0.033 |
| | | VTE | 0.191 | -0.014 | 0.077 | 0.195 | -0.006 | 0.040 | 0.199 | -0.002 | 0.031 |
| $\boldsymbol{B}$ | 0.80 | EbEE | 0.799 | -0.005 | 0.042 | 0.800 | -0.002 | 0.023 | 0.799 | -0.001 | 0.012 |
| | | VTE | 0.798 | -0.001 | 0.035 | 0.799 | -0.001 | 0.018 | 0.800 | -0.001 | 0.015 |
| | 0.90 | EbEE | 0.899 | -0.004 | 0.029 | 0.898 | -0.002 | 0.013 | 0.899 | -0.001 | 0.010 |
| | | VTE | 0.900 | -0.003 | 0.023 | 0.900 | -0.001 | 0.010 | 0.900 | -0.001 | 0.008 |
| $\boldsymbol{C}$ | 0.15 | EbEE | 0.150 | -0.002 | 0.023 | 0.150 | 0.000 | 0.010 | 0.150 | 0.000 | 0.008 |
| | | VTE | 0.149 | -0.001 | 0.015 | 0.149 | 0.000 | 0.009 | 0.150 | 0.000 | 0.007 |
| | 0.10 | EbEE | 0.105 | 0.008 | 0.053 | 0.101 | 0.004 | 0.032 | 0.102 | 0.003 | 0.025 |
| | | VTE | 0.101 | 0.000 | 0.027 | 0.100 | 0.000 | 0.016 | 0.100 | 0.000 | 0.012 |
| | 0.05 | EbEE | 0.050 | 0.000 | 0.025 | 0.050 | 0.000 | 0.023 | 0.051 | 0.000 | 0.010 |
| | | VTE | 0.051 | 0.001 | 0.016 | 0.050 | 0.000 | 0.008 | 0.050 | 0.000 | 0.007 |
| | 0.20 | EbEE | 0.195 | -0.011 | 0.039 | 0.200 | -0.003 | 0.023 | 0.199 | -0.002 | 0.017 |
| | | VTE | 0.199 | -0.001 | 0.025 | 0.200 | 0.001 | 0.013 | 0.200 | 0.000 | 0.011 |

RMSE is the Root Mean Square Error.

the remaining parameters of the intercept matrix in the second step. The strong consistency of the EbEE is showed under the assumptions that are weaker than the ones for the VTE of the same model. Under the mixing-conditions, the asymptotic distribution of the estimators of the parameters is normal. The main motivation for using the EbE method in application is the important gains in computational time and our experiments show that the reduction of computational time compared to the VT estimation can be effective.

# 6 Proofs

**Proof of Theorem 1.**

Let

$$Q_n^{(k)}(\boldsymbol{\theta}^{(k)}) = \frac{1}{n}\sum_{t=1}^{n} \ell_{kt}(\boldsymbol{\theta}^{(k)}), \quad \ell_{kt}(\boldsymbol{\theta}^{(k)}) = \log \sigma_{kt}^2(\boldsymbol{\theta}^{(k)}) + \frac{\varepsilon_{kt}^2}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}.$$

To prove the theorem of the consistency of the EbEE, the following results have to be shown

i) $\lim_{n\to\infty} \sup_{\boldsymbol{\theta}^{(k)}\in\boldsymbol{\Theta}^{(k)}} \left| Q_n^{(k)}(\boldsymbol{\theta}^{(k)}) - \widetilde{Q}_n^{(k)}(\boldsymbol{\theta}^{(k)}) \right| = 0$ a.s.

ii) $\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)}) = \sigma_{kt}^2(\boldsymbol{\theta}^{(k)})$ a.s. iff $\boldsymbol{\theta}_0^{(k)} = \boldsymbol{\theta}^{(k)}$.

iii) $E\ell_{kt}(\boldsymbol{\theta}_0^{(k)}) < \infty$ and if $\boldsymbol{\theta}^{(k)} \neq \boldsymbol{\theta}_0^{(k)}$ then $E\ell_{kt}(\boldsymbol{\theta}_0^{(k)}) < E\ell_{kt}(\boldsymbol{\theta}^{(k)})$.

iv) There exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}^{(k)})$ of any $\boldsymbol{\theta}^{(k)} \neq \boldsymbol{\theta}_0^{(k)}$ such that

$$\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}^*\in\mathcal{V}(\boldsymbol{\theta}^{(k)})} \widetilde{Q}_n^{(k)}(\boldsymbol{\theta}^*) > \limsup_{n\to\infty} \widetilde{Q}_n^{(k)}(\boldsymbol{\theta}_0^{(k)}), \text{ a.s.}$$

The condition $\sum_{k=1}^{m} b_k^2 < 1$ of Assumption **A5** and the compactness of $\boldsymbol{\Theta}^{(k)}$ imply that

$$\sup_{\boldsymbol{\theta}^{(k)}\in\boldsymbol{\Theta}^{(k)}} b_k^2 < 1. \tag{20}$$

By a simple recursion, we have

$$\sup_{\boldsymbol{\theta}^{(k)}\in\boldsymbol{\Theta}^{(k)}} \left| \widetilde{\sigma}_{kt}^2(\boldsymbol{\theta}^{(k)}) - \sigma_{kt}^2(\boldsymbol{\theta}^{(k)}) \right| = \sup_{\boldsymbol{\theta}^{(k)}\in\boldsymbol{\Theta}^{(k)}} (b_k^2)^t \left| \widetilde{\sigma}_{k0}^2(\boldsymbol{\theta}^{(k)}) - \sigma_{k0}^2(\boldsymbol{\theta}^{(k)}) \right| < K\rho^t \text{ a.s.}$$

where, here and the sequel of the paper, $K$ and $\rho$ denote generic constants whose the exact values are not important and $0 < \rho < 1$.

The proofs of i), iii) and i) are very similar to the ones given in Francq and Zakoïan (2010) for the standard GARCH without covariates and are omitted. Here I only thus show the point ii).

Let $L$ be the back-shift operator, *i.e.* $L(u_t) = u_{t-1}$. By Assumption **A5**, the polynomial $1 - b_k^2 L$ is invertible for any $\boldsymbol{\theta}^{(k)} \in \boldsymbol{\Theta}^{(k)}$. Assume that $\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)}) = \sigma_{kt}^2(\boldsymbol{\theta}^{(k)})$ a.s. We have

$$\begin{aligned}
&(1-(b_k^0)^2L)^{-1}\left(\sum_{\ell=1}^{m} a_{k\ell}^0 \varepsilon_{\ell,t-1}\right)^2 - (1-b_k^2L)^{-1}\left(\sum_{\ell=1}^{m} a_{k\ell}\varepsilon_{\ell,t-1}\right)^2 \\
&+ (1-(b_k^0)^2L)^{-1}\left(\sum_{s=1}^{r} c_{k\ell}^0 x_{s,t-1}\right)^2 - (1-b_k^2L)^{-1}\left(\sum_{s=1}^{r} c_{ks}x_{s,t-1}\right)^2 \\
=&(1-b_k^2)^{-1}\omega_{kk} - (1-(b_k^0)^2)^{-1}\omega_{kk}^0 \qquad \text{a.s.}
\end{aligned}$$

Then

$$\sum_{i=0}^{\infty}\sum_{\ell,\ell'=1}^{m} \overline{a}_{i\ell\ell'}^{(k)} \varepsilon_{\ell,t-i-1}\varepsilon_{\ell',t-i-1} + \sum_{j=0}^{\infty}\sum_{s,s'=1}^{r} \overline{c}_{jss'}^{(k)} x_{s,t-j-1}x_{s',t-j-1} = c \quad \text{a.s.} \tag{21}$$

where $\overline{a}_{i\ell\ell'}^{(k)} = (b_k^0)^{2i} a_{k\ell}^0 a_{k\ell'}^0 - (b_k)^{2i} a_{k\ell}a_{k\ell'}$, $\overline{c}_{jss'}^{(k)} = (b_k^0)^{2i} c_{ks}^0 c_{ks'}^0 - (b_k)^{2i} c_{ks}c_{ks'}$ and $c = (1 - (b_k^0)^2)^{-1}\omega_{kk} - (1-b_k^2)^{-1}\omega_{kk}^0$. If $b_k^0 \neq b_k$ or there exists $\ell^*$ such that $a_{k\ell^*}^0 \neq a_{k\ell^*}$ then $\varepsilon_{\ell^*,t-1}^2$ is a linear combination of the $x_{s,u}x_{s',u}, s, s' = 1,\dots,r, u < t$, the $\varepsilon_{\ell,v}\varepsilon_{\ell',v}, \ell, \ell' = 1,\dots,m, v < t-1$ and the $\varepsilon_{\ell,t-1}\varepsilon_{\ell',t-1}, (\ell,\ell') \neq (\ell^*,\ell^*)$ which is impossible by Assumption **A7**. Therefore $b_k^0 = b_k$ and $a_{k\ell}^0 = a_{k\ell}$ for all $\ell = 1,\dots,m$ and (21) becomes

$$\sum_{j=0}^{\infty}\sum_{s,s'=1}^{r} \overline{c}_{jss'}^{(k)} x_{s,t-j-1}x_{s',t-j-1} = c \quad \text{a.s.}$$

Similarly, if there exists $s^*$ such that $c_{ks^*}^0 \neq c_{ks^*}$, $x_{s^*,t-1}^2$ is a linear combination of $x_{u,t-j}x_{u',t-j}, u, u' = 1,\dots,r, j > 1$ and $x_{s,t-1}x_{s',t-1}, (s,s') \neq (s^*,s^*)$ which contradicts Assumption **A8**. Therefore we have $c_{ks}^0 = c_{ks}$ for all $s = 1,\dots,r$. Hence ii) is proved. $\square$

For the proof the asymptotic distribution in the Theorem 2, we need the following lemmas

**Lemma 1** *Under assumptions of Theorem 2, there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$*

*such that*

$$E\left\{\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{1}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}\frac{\partial\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\theta}^{(k)}}\right\|^{4(1+1/\delta)}\right\}<\infty,\quad \textit{for some } \delta>0, \tag{22}$$

$$E\left\{\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left\|\frac{1}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}\frac{\partial^2\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial\boldsymbol{\theta}^{(k)}\partial\boldsymbol{\theta}^{(k)'}}\right\|^{2(1+1/\delta)}\right\}<\infty,\quad \textit{for some } \delta>0, \tag{23}$$

$$E\left\{\sup_{\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})}\left|\frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}\right|^{s}\right\}<\infty,\quad \textit{for any } s>0. \tag{24}$$

**Proof of Lemma 1**

Iteratively using the volatility equation in (4), we obtain

$$\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})=\sum_{j=0}^{\infty}b_k^{2j}\left\{\omega_{kk}+\sum_{\ell,\ell'=1}^{m}a_{k\ell}a_{k\ell'}\varepsilon_{\ell,t-j-1}\varepsilon_{\ell',t-j-1}+\sum_{s,s'=1}^{r}c_{ks}c_{ks'}x_{s,t-j-1}x_{s',t-j-1}\right\}. \tag{25}$$

Derive (25) with respect to $\boldsymbol{\theta}^{(k)}$, we get

$$\begin{aligned}
\frac{\partial\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial\omega_{kk}}&=\sum_{j=0}^{\infty}b_k^{2j}=\frac{1}{1-b_k^2},\\
\frac{\partial\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial a_{k\ell}}&=2\sum_{j=0}^{\infty}b_k^{2j}\sum_{\ell'=1}^{m}a_{k\ell'}\varepsilon_{\ell,t-j-1}\varepsilon_{\ell',t-j-1},\\
\frac{\partial\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial b_k^2}&=\sum_{j=1}^{\infty}jb_k^{2(j-1)}\left\{\omega_{kk}+\sum_{\ell,\ell'=1}^{m}a_{k\ell}a_{k\ell'}\varepsilon_{\ell,t-j-1}\varepsilon_{\ell',t-j-1}+\sum_{s,s'=1}^{r}c_{ks}c_{ks'}x_{s,t-j-1}x_{s',t-j-1}\right\},\\
\frac{\partial\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial c_{ks}}&=2\sum_{j=0}^{\infty}b_k^{2j}\sum_{s'=1}^{m}c_{ks'}x_{s,t-j-1}x_{s',t-j-1}
\end{aligned}.$$

Silmilar expressions hold for the second order derivatives. Noting that

$$\underline{\omega}_{kk}:=\inf_{\boldsymbol{\theta}^{(k)}\in\boldsymbol{\Theta}^{(k)}}\sigma_{kt}^2>0.$$

Using the moment conditions **A10** and (20), we obtain (22) and (23).

The moment condition (24) will be showed even if some components of $\underline{\boldsymbol{a}}_k^0$ or $\underline{\boldsymbol{c}}_k^0$ are zero. Indeed, there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ such that for all $\boldsymbol{\theta}^{(k)}\in\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$

$$\frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}\le K+K\sum_{j=0}^{\infty}\frac{(b_k^0)^{2j}}{b_k^{2j}}\left\{\frac{\left(\sum_{\substack{\ell=1\\ a_{k\ell}^0\neq0}}^{m}\frac{a_{k\ell}}{\sqrt{\omega_{kk}}}\varepsilon_{\ell,t-j-1}\right)^2}{1+\left(\sum_{\substack{\ell=1\\ a_{k\ell}^0\neq0}}^{m}\frac{a_{k\ell}}{\sqrt{\omega_{kk}}}\varepsilon_{\ell,t-j-1}\right)^2}+\frac{\left(\sum_{\substack{s=1\\ c_{ks}^0\neq0}}^{r}\frac{c_{ks}}{\sqrt{\omega_{kk}}}x_{s,t-j-1}\right)^2}{1+\left(\sum_{\substack{s=1\\ c_{ks}^0\neq0}}^{r}\frac{c_{ks}}{\sqrt{\omega_{kk}}}x_{s,t-j-1}\right)^2}\right\}.$$

For all $\delta > 0$, there exists $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ such that $b_k^0 \leq (1+\delta)b_k$ for all $\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})$. It follows that for all $\delta > 0$ and $u \in (0,1)$, using the inequality $z/(1+z) \leq z^u$ for all $z \geq 0$, there exists $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ such that

$$\sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \leq K + K \sum_{j=0}^{\infty} (1+\delta)^j \rho^{ju} \left\{ \left( \sum_{\substack{\ell=1 \\ a_{k\ell}^0 \neq 0}}^{m} \varepsilon_{\ell,t-j-1} \right)^{2u} + \left( \sum_{\substack{s=1 \\ c_{ks}^0 \neq 0}}^{r} x_{s,t-j-1} \right)^{2u} \right\}.$$

Denote by $\|.\|_d$ the $L^d$ norm, for $d \geq 1$, on the space of real random variables. Using the Minskowski inequality and choosing $u$ such that $E\|\varepsilon_1\|^{2us} < \infty$ and $E\|x_1\|^{2us} < \infty$ and choosing, for instance, $\delta = \dfrac{1-\rho^u}{2\rho^u}$ and by Assumption **A6**, we have

$$\begin{aligned} \left\| \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \right\|_s &\leq K + K \sum_{j=0}^{\infty} (1+\delta)^j \rho^{ju} \left\{ \sum_{\substack{\ell=1 \\ a_{k\ell}^0 \neq 0}}^{m} \|\varepsilon_{\ell,t-j-1}\|_{2us}^{2u} + \sum_{\substack{s=1 \\ c_{ks}^0 \neq 0}}^{r} \|x_{s,t-j-1}\|_{2us}^{2u} \right\} \\ &< \infty. \end{aligned}$$

(24) is thus obtained. □

**Lemma 2** *Under assumptions of Theorem 2, for all $t$, we have*

*i)* $E \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \dfrac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} \right\| < \infty$, *for some neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$.*

*ii)* $\dfrac{1}{\sqrt{n}} \sum_{t=1}^{n} \dfrac{\partial^2 \ell_{kt}(\bar{\boldsymbol{\theta}}_n^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} \to \boldsymbol{J}_{kk}$, *a.s. for any $\bar{\boldsymbol{\theta}}_n^{(k)}$ between $\hat{\boldsymbol{\theta}}_n^{(k)}$ and $\boldsymbol{\theta}_0^{(k)}$.*

*iii)* $\boldsymbol{J}_{kk}$ *is non singular.*

**Proof of Lemma 2**

The derivatives of $\ell_{kt}(\boldsymbol{\theta}^{(k)})$ is given by

$$\frac{\partial \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} = \left\{ 1 - \frac{\varepsilon_{kt}^2}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \right\} \left\{ \frac{1}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \frac{\partial \sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} \right\} \tag{26}$$

and

$$\begin{aligned} \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} &= \left\{ 1 - \frac{\varepsilon_{kt}^2}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \right\} \left\{ \frac{1}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \frac{\partial^2 \sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} \right\} \\ &\quad + \left\{ 2 \frac{\varepsilon_{kt}^2}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} - 1 \right\} \left\{ \frac{1}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \frac{\partial \sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} \right\} \left\{ \frac{1}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \frac{\partial \sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)'}} \right\}. \end{aligned} \tag{27}$$

Using the triangle inequality, i) is obtained by showing the existence of the expectations of the two terms in the right-hand side of (27). Let us consider the first one. We have, by the Holder inequality,

$$E \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \left\{ 1 - \frac{\varepsilon_{kt}^2}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \right\} \left\{ \frac{1}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \frac{\partial^2 \sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} \right\} \right\|$$
$$\leq \left\{ 1 + \|\eta_{kt}^{*2}\|_{2(\delta+1)} \left\| \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \frac{\sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \right\|_2 \right\} \left\| \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{1}{\sigma_{kt}^2(\boldsymbol{\theta}^{(k)})} \frac{\partial^2 \sigma_{kt}^2(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} \right\| \right\|_{2(1+1/\delta)}$$

which is finite by (23) and (24).The second product in the right-hand side of (27) can be handled similarly using (22).

To prove ii), by Exercise 7.9 in Francq and Zakoïan (2010), it will be sufficient to establish that for any $\epsilon > 0$, there exists a neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ of $\boldsymbol{\theta}_0^{(k)}$ such that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} - \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} \right\| \leq \epsilon \qquad a.s. \tag{28}$$

By the ergodic theorem, the limit in the left-hand side is equal to

$$E \sup_{\boldsymbol{\theta}^{(k)} \in \mathcal{V}(\boldsymbol{\theta}_0^{(k)})} \left\| \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} - \frac{\partial^2 \ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}} \right\|$$

that is finite by i). This expectation tends to zero when the neighborhood $\mathcal{V}(\boldsymbol{\theta}_0^{(k)})$ shrinks to singleton $\{\boldsymbol{\theta}_0^{(k)}\}$. The point ii) is thus proved.

We now turn show the invertibility of $\boldsymbol{J}_{kk}$. Assume that $\boldsymbol{J}_{kk}$ is singular. Then there exists a vector $\boldsymbol{\pi} \in \mathbb{R}^{2+m+r}$ such that $\boldsymbol{\pi}' \dfrac{\partial \sigma_{kt}^2(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} = 0$ a.s. It follows that

$$\pi_1 + 2\left(\sum_{\ell=1}^{m} a_{k\ell}^0 \varepsilon_{\ell,t-1}\right) \boldsymbol{\pi}' \begin{pmatrix} 0 \\ \boldsymbol{\varepsilon}_{t-1} \\ \mathbf{0}_{(1+r)\times 1} \end{pmatrix} + \pi_{2+m} \sigma_{k,t-1}^2(\boldsymbol{\theta}_0^{(k)})$$
$$+2\left(\sum_{s=1}^{r} c_{ks}^0 x_{s,t-1}\right) \boldsymbol{\pi}' \begin{pmatrix} \mathbf{0}_{(2+m)\times 1} \\ \boldsymbol{x}_{t-1} \end{pmatrix} + b_k^{02} \boldsymbol{\pi}' \frac{\partial \sigma_{k,t-1}^2(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} = 0, \qquad \text{a.s.}$$

Note that the last term in the left-hand side of this equation is equal to zero by the stationarity. We thus have

$$\pi_1 + 2\sum_{\ell,\ell'=1}^{m} a_{k\ell}^0 \pi_{1+\ell'} \varepsilon_{\ell,t-1} \varepsilon_{\ell',t-1} + \pi_{2+m} \sigma_{k,t-1}^2(\boldsymbol{\theta}_0^{(k)}) + 2\sum_{s,s'=1}^{r} c_{ks}^0 \pi_{2+m+s'} x_{s,t-1} x_{s',t-1} = 0 \text{ a.s.}$$

By the similar arguments used to show ii) in the proof of Theorem 1, we can conclude that $\boldsymbol{\pi} = 0$. $\square$

**Proof of Theorem 2**

Since $\widehat{\boldsymbol{\theta}}_n^{(k)}$ strongly converges to $\boldsymbol{\theta}_0^{(k)}$ which belongs to the interior of the parameter space $\boldsymbol{\Theta}^{(k)}$, the derivative of the criterion $\widetilde{Q}_n^{(k)}$ is equal to zero at $\widehat{\boldsymbol{\theta}}_n^{(k)}$. Let $\boldsymbol{J}_{kkn}(\boldsymbol{\theta}^{(k)}) = \dfrac{1}{n}\sum_{t=1}^{n} \dfrac{\partial^2 \ell_{kt}(\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^{(k)} \partial \boldsymbol{\theta}^{(k)'}}$. Applying a Taylor expansion for $Q_n^{(k)}$ at $\boldsymbol{\theta}_0^{(k)}$ and the mean-value theorem gives

$$0 = \frac{1}{n}\sum_{t=1}^{n} \frac{\ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}} + \boldsymbol{J}_{kkn}\left(\bar{\boldsymbol{\theta}}_n^{(k)}\right)\left(\widehat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}\right), \tag{29}$$

where $\bar{\boldsymbol{\theta}}_n^{(k)}$ is between $\widehat{\boldsymbol{\theta}}_n^{(k)}$ and $\boldsymbol{\theta}_0^{(k)}$. By the points ii) and iii) in Lemma 2 and the consistency of $\widehat{\boldsymbol{\theta}}_n^{(k)}$, the matrix $\boldsymbol{J}_{kkn}\left(\bar{\boldsymbol{\theta}}_n^{(k)}\right)$ is a.s. invertible for sufficiently large $n$. Hence multiplying by $\sqrt{n}$ and solving for $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}\right)$ gives

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n^{(k)} - \boldsymbol{\theta}_0^{(k)}\right) \overset{o_p(1)}{=} -\boldsymbol{J}_{kkn}^{-1}\left(\bar{\boldsymbol{\theta}}_n^{(k)}\right)\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{\ell_{kt}(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}. \tag{30}$$

Let $\dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}) = \left(\dfrac{\partial \ell_{1t}(\boldsymbol{\theta}^{(1)})}{\partial \boldsymbol{\theta}^{(1)'}}, \dots, \dfrac{\partial \ell_{mt}(\boldsymbol{\theta}^{(m)})}{\partial \boldsymbol{\theta}^{(m)'}}\right)'$. Collecting all these Taylor expansions, we have

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \overset{o_p(1)}{=} \boldsymbol{J}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0).$$

Note that $\boldsymbol{\eta}_t^* = \boldsymbol{D}_{0t}^{-1}\boldsymbol{H}_{0t}^{1/2}\boldsymbol{\eta}_t$. From (26), we then have

$$\begin{aligned}\dot{\boldsymbol{\ell}}_t(\boldsymbol{\theta}_0) &= -\boldsymbol{\Delta}_t T_m vec\left(\boldsymbol{\eta}_t^*\boldsymbol{\eta}_t^{*'} - \boldsymbol{D}_{0t}^{-1}\boldsymbol{H}_{0t}\boldsymbol{D}_{0t}^{-1}\right)\\ &= -\boldsymbol{\Delta}_t T_m\left(\boldsymbol{D}_{0t}^{-1}\boldsymbol{H}_{0t}^{1/2}\right)^{\otimes 2} vec\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' - I_m\right).\end{aligned}$$

Then, we get

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \overset{o_p(1)}{=} -\boldsymbol{J}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{\Delta}_t T_m\left(\boldsymbol{D}_{0t}^{-1}\boldsymbol{H}_{0t}^{1/2}\right)^{\otimes 2} vec\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' - I_m\right). \tag{31}$$

We now introduce the martingale difference

$$\boldsymbol{\nu}_t = vec(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') - vec(\boldsymbol{H}_{0t}) = \left(\boldsymbol{H}_{0t}^{1/2} \otimes \boldsymbol{H}_{0t}^{1/2}\right) vec(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' - I_m).$$

In the representation of $vec(\boldsymbol{H}_{0t})$ obtained from (2), we replace $vec(\boldsymbol{H}_{0t})$ by $vec(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')-\boldsymbol{\nu}_t$. Then, we get

$$\begin{aligned} vec\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' - E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t')\right) = \left(\boldsymbol{A}_0^{\otimes 2} + \boldsymbol{B}_0^{\otimes 2}\right) vec\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}' - E(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}')\right) \\ + \boldsymbol{C}_0^{\otimes 2} vec\left(\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}' - E\boldsymbol{x}_{t-1}\boldsymbol{x}_{t-1}'\right) + \left(\boldsymbol{\nu}_t - \boldsymbol{B}_0^{\otimes 2}\boldsymbol{\nu}_{t-1}\right). \end{aligned}$$

Note that under assumption **A5**, the matrix $I_{m^2} - \boldsymbol{A}_0^{\otimes 2} - \boldsymbol{B}_0^{\otimes 2}$ is inversible. Taking the average of the two sides of the equality for $t = 1, \dots, n$ gives

$$\begin{aligned} \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon},n} - \boldsymbol{\gamma}_{\boldsymbol{\varepsilon},0} =& L_m(I_{m^2} - \boldsymbol{A}_0^{\otimes 2} - \boldsymbol{B}_0^{\otimes 2})^{-1}(I_{m^2} - \boldsymbol{B}_0^{\otimes 2})\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\nu}_t \\ &+ L_m(I_{m^2} - \boldsymbol{A}_0^{\otimes 2} - \boldsymbol{B}_0^{\otimes 2})^{-1}\boldsymbol{C}_0^{\otimes 2} D_r\left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{x},n} - \boldsymbol{\gamma}_{\boldsymbol{x},0}\right) + o_p(1), \qquad a.s. \end{aligned}$$

We then have

$$\begin{pmatrix} \sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \\ \sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}n} - \boldsymbol{\gamma}_{\boldsymbol{\varepsilon}}\right) \\ \sqrt{n}\left(\widehat{\boldsymbol{\gamma}}_{\boldsymbol{x}n} - \boldsymbol{\gamma}_{\boldsymbol{x}}\right) \end{pmatrix} \overset{o_p(1)}{=} \begin{pmatrix} -\boldsymbol{J}^{-1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{N}_1 & \boldsymbol{N}_2 \\ \boldsymbol{0} & \boldsymbol{0} & I_{r(r+1)/2} \end{pmatrix} \frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^{n} \boldsymbol{\Upsilon}_{0t} vec\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' - I_m\right) \\ \sum_{t=1}^{n} vech(\boldsymbol{x}_t\boldsymbol{x}_t' - E(\boldsymbol{x}_t\boldsymbol{x}_t)) \end{pmatrix}.$$

The arguments for establishing the limiting distribution of $\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^{n} \boldsymbol{\Upsilon}_{0t} vec\left(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' - I_m\right) \\ \sum_{t=1}^{n} vech(\boldsymbol{x}_t\boldsymbol{x}_t' - E(\boldsymbol{x}_t\boldsymbol{x}_t)) \end{pmatrix}$ being very similar to Lemma 6 in Thieu (2016), I just give a sketch of proof. For any $s > 0$, the $k$-th individual volatility in (4) can be written $\sigma_{kt}^2 = \underline{\sigma}_{kts}^2 + \overline{\sigma}_{kts}^2$, where

$$\begin{aligned} \underline{\sigma}_{kts}^2 &= \sum_{j=0}^{s} b_k^{2j}\left\{\omega_{kk} + \left(\sum_{\ell=1}^{m} a_{k\ell}\varepsilon_{\ell,t-j-1}\right)^2 + \left(\sum_{s=1}^{r} c_{ks}x_{s,t-j-1}\right)^2\right\} \\ \overline{\sigma}_{kts}^2 &= \sum_{j=s+1}^{\infty} b_k^{2j}\left\{\omega_{kk} + \left(\sum_{\ell=1}^{m} a_{k\ell}\varepsilon_{\ell,t-j-1}\right)^2 + \left(\sum_{s=1}^{r} c_{ks}x_{s,t-j-1}\right)^2\right\}. \end{aligned}$$

Then we can write $\boldsymbol{\Upsilon}_{0t} vec(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' - I_m) = \underline{\boldsymbol{Y}}_{t,s} + \boldsymbol{Y}_{t,s}^*$, where

$$\underline{\boldsymbol{Y}}_{t,s} = \begin{pmatrix} \underline{\boldsymbol{\Delta}}_{ts} T_m \left(\underline{\boldsymbol{D}}_{0ts}^{-1}\underline{\boldsymbol{H}}_{0t,s}^{1/2}\right) \otimes \left(\underline{\boldsymbol{D}}_{0ts}^{-1}\underline{\boldsymbol{H}}_{0t,s}^{1/2}\right) \\ \underline{\boldsymbol{H}}_{0t,s}^{1/2} \otimes \underline{\boldsymbol{H}}_{0t,s}^{1/2} \end{pmatrix} vec(\boldsymbol{\eta}_t\boldsymbol{\eta}_t' - I_m),$$

with $\underline{\boldsymbol{D}}_{0ts} = diag(\underline{\sigma}_{1ts}, \dots, \underline{\sigma}_{mts})$ and $\underline{\boldsymbol{\Delta}}_{ts} = diag(\underline{\Delta}_{1ts}, \dots \underline{\Delta}_{mts})$, $\underline{\Delta}_{kts} = \dfrac{1}{\underline{\sigma}_{kts}^2}\dfrac{\partial \underline{\sigma}_{kts}^2(\boldsymbol{\theta}_0^{(k)})}{\partial \boldsymbol{\theta}^{(k)}}$, for $k = 1, \dots, m$ and $\boldsymbol{Y}_{t,s}^*$ is stationary and centered process satisfying

$$\lim_{s\to\infty}\limsup_{n\to\infty} P\left(\left\|n^{-1/2}\sum_{t=1}^{n}\boldsymbol{Y}_{t,s}^*\right\| > \epsilon\right) = 0.$$

Under Assumption **A11** and using the same argument as in the reference, we have for some $\nu$ and $\delta > 0$

$$\left\| \left( \underline{\boldsymbol{H}}_{0t,s}^{1/2} \otimes \underline{\boldsymbol{H}}_{0t,s}^{1/2} \right) vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) \right\|_{2+\nu} < \infty.$$

Note that $\boldsymbol{D}_{0t}^{-1} \boldsymbol{H}_{0t} \boldsymbol{D}_{0t}^{-1}$ is the conditional correlation matrix of $\boldsymbol{\varepsilon}_t$. Because $\boldsymbol{D}_{0t}^{-1} \boldsymbol{H}_{0t} \boldsymbol{D}_{0t}^{-1} = \left( \boldsymbol{D}_{0t}^{-1} \boldsymbol{H}_{0t}^{1/2} \right) \left( \boldsymbol{D}_{0t}^{-1} \boldsymbol{H}_{0t}^{1/2} \right)'$, all the elements of the matrix $\boldsymbol{D}_{0t}^{-1} \boldsymbol{H}_{0t}^{1/2}$ are smaller than one. It is thus easy to show that all the ones of the matrix $\left( \underline{\boldsymbol{D}}_{0t,s}^{-1} \underline{\boldsymbol{H}}_{0t,s}^{1/2} \right) \otimes \left( \underline{\boldsymbol{D}}_{0t,s}^{-1} \underline{\boldsymbol{H}}_{0t,s}^{1/2} \right)$ are also smaller than one. Using the Holder inequality, we then have

$$\begin{aligned}
&\left\| \underline{\boldsymbol{\Delta}}_{ts} T_m \left( \underline{\boldsymbol{D}}_{0t,s}^{-1} \underline{\boldsymbol{H}}_{0t,s}^{1/2} \right) \otimes \left( \underline{\boldsymbol{D}}_{0t,s}^{-1} \underline{\boldsymbol{H}}_{0t,s}^{1/2} \right) vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) \right\|_{2+\nu} \\
&\leq \| \underline{\boldsymbol{\Delta}}_{ts} \|_{(2+\nu)(1+1/\delta)} \left\| T_m \left( \underline{\boldsymbol{D}}_{0t,s}^{-1} \underline{\boldsymbol{H}}_{0t,s}^{1/2} \right) \otimes \left( \underline{\boldsymbol{D}}_{0t,s}^{-1} \underline{\boldsymbol{H}}_{0t,s}^{1/2} \right) vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) \right\|_{(2+\nu)(1+\delta)} \\
&\leq K \| \underline{\boldsymbol{\Delta}}_{ts} \|_{(2+\nu)(1+1/\delta)} \| vec(\boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m) \|_{(2+\nu)(1+\delta)} \\
&< \infty,
\end{aligned}$$

where the last inequality follows from (20) and Assumption **A11**. It implies that $\| \underline{\boldsymbol{Y}}_{t,s} \|_{2+\nu} < \infty$. The process $(\underline{\boldsymbol{Y}}_{t,s})_t$, for $s$ fixed, is thus strongly mixing under Assumption **A11**. Applying the central limit theorem of Herrndorf (1984), we get

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{t=1}^n \boldsymbol{\Upsilon}_{0t} vec\left( \boldsymbol{\eta}_t \boldsymbol{\eta}_t' - I_m \right) \\ \sum_{t=1}^n vech(\boldsymbol{x}_t \boldsymbol{x}_t' - E(\boldsymbol{x}_t \boldsymbol{x}_t)) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\Sigma}\right).$$

The asymptotic distribution of Theorem 2 thus follows from the Slutzky theorem. □

**The proof of Theorem 3**

The strong consistency of $\widehat{\boldsymbol{\xi}}_n$ is obviously obtained.

Now we turn to its asymptotic normality.

Using the elementary relation $vec(ABC) = (C' \otimes A) vec(B)$, we have

$$\begin{aligned}
vec(\widehat{\boldsymbol{A}}_n \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n} \widehat{\boldsymbol{A}}_n' - \boldsymbol{A}_0 \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}0} \boldsymbol{A}_0') =& vec\left( \widehat{\boldsymbol{A}}_n \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n} (\widehat{\boldsymbol{A}}_n' - \boldsymbol{A}_0') \right) + vec\left( \widehat{\boldsymbol{A}}_n (\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n} - \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \boldsymbol{A}_0' \right) \\
&+ vec\left( (\widehat{\boldsymbol{A}}_n - \boldsymbol{A}_0) \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \boldsymbol{A}_0' \right) \\
=& \left\{ I_m \otimes (\widehat{\boldsymbol{A}}_n \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n}) + ((\boldsymbol{A}_0 \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \otimes I_m) M_{mm} \right\} vec(\widehat{\boldsymbol{A}}_n' - \boldsymbol{A}_0') \\
&+ (\boldsymbol{A}_0 \otimes \widehat{\boldsymbol{A}}_n) D_m vech(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n} - \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}).
\end{aligned}$$

Similary, we also get

$$vec(\widehat{\boldsymbol{B}}_n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n}\widehat{\boldsymbol{B}}_n - \boldsymbol{B}_0\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}\boldsymbol{B}_0) = \left\{I_m \otimes (\widehat{\boldsymbol{B}}_n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n}) + ((\boldsymbol{B}_0\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) \otimes I_m)\right\} vec(\widehat{\boldsymbol{B}}_n - \boldsymbol{B}_0)$$
$$+ (\boldsymbol{B}_0 \otimes \widehat{\boldsymbol{B}}_n)D_m vech(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n} - \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$$

and

$$vec(\widehat{\boldsymbol{C}}_n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}n}\widehat{\boldsymbol{C}}'_n - \boldsymbol{C}_0\boldsymbol{\Sigma}_{\boldsymbol{x}0}\boldsymbol{C}'_0) = \left\{I_m \otimes (\widehat{\boldsymbol{C}}_n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}n}) + ((\boldsymbol{C}_0\boldsymbol{\Sigma}_{\boldsymbol{x}}) \otimes I_m)M_{mr}\right\} vec(\widehat{\boldsymbol{C}}'_n - \boldsymbol{C}'_0)$$
$$+ (\boldsymbol{C}_0 \otimes \widehat{\boldsymbol{C}}_n)D_r vech(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}n} - \boldsymbol{\Sigma}_{\boldsymbol{x}}).$$

Let

$$\begin{aligned}
\widehat{\boldsymbol{A}}^*_n &= -P_m\left\{I_m \otimes (\widehat{\boldsymbol{A}}_n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n}) + ((\boldsymbol{A}_0\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}0}) \otimes I_m)M_{mm}\right\},\\
\widehat{\boldsymbol{B}}^*_n &= -P_m\left\{I_m \otimes (\widehat{\boldsymbol{B}}_n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}n}) + (\boldsymbol{B}_0\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}0}) \otimes I_m\right\},\\
\widehat{\boldsymbol{C}}^*_n &= -P_m\left\{I_m \otimes (\widehat{\boldsymbol{C}}_n\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{x}n}) + ((\boldsymbol{C}_0\boldsymbol{\Sigma}_{\boldsymbol{x}0}) \otimes I_m)M_{mr}\right\},\\
\widehat{\boldsymbol{E}}^*_n &= P_m(I_{m^2} - \boldsymbol{B}_0 \otimes \widehat{\boldsymbol{B}}_n - \boldsymbol{A}_0 \otimes \widehat{\boldsymbol{A}}_n)D_m, \quad \widehat{\boldsymbol{X}}^*_n = -P_m(\boldsymbol{C}_0 \otimes \widehat{\boldsymbol{C}}_n)D_r.
\end{aligned}$$

We then have

$$\sqrt{n}(\widehat{\underline{\boldsymbol{\omega}}}_n - \underline{\boldsymbol{\omega}}_0) = \begin{pmatrix} \widehat{\boldsymbol{A}}^*_n & \widehat{\boldsymbol{B}}^*_n & \widehat{\boldsymbol{C}}^*_n & \widehat{\boldsymbol{E}}^*_n & \widehat{\boldsymbol{X}}^*_n \end{pmatrix} \sqrt{n} \begin{pmatrix} \widehat{\underline{\boldsymbol{a}}}_n - \underline{\boldsymbol{a}}_0 \\ \widehat{\underline{\boldsymbol{b}}}_n - \underline{\boldsymbol{b}}_0 \\ \widehat{\underline{\boldsymbol{c}}}_n - \underline{\boldsymbol{c}}_0 \\ \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}n} - \boldsymbol{\gamma}_{\boldsymbol{\varepsilon}0} \\ \widehat{\boldsymbol{\gamma}}_{\boldsymbol{x}n} - \boldsymbol{\gamma}_{\boldsymbol{x}0} \end{pmatrix}, \tag{32}$$

where $\widehat{\underline{\boldsymbol{a}}}_n = vec(\widehat{\boldsymbol{A}}'_n)$, $\widehat{\underline{\boldsymbol{b}}}_n = vec(\widehat{\boldsymbol{B}}_n)$ and $\widehat{\underline{\boldsymbol{c}}}_n = vec(\widehat{\boldsymbol{C}}'_n)$ are the estimators of $\underline{\boldsymbol{a}}_0 = vec(\boldsymbol{A}'_0)$, $\underline{\boldsymbol{b}}_0 = vec(\boldsymbol{B}_0)$ and $\underline{\boldsymbol{c}}_0 = vec(\boldsymbol{C}'_0)$ respectively. Note that

$$\begin{pmatrix} \widehat{\underline{\boldsymbol{a}}}_n - \underline{\boldsymbol{a}}_0 \\ \widehat{\underline{\boldsymbol{b}}}_n - \underline{\boldsymbol{b}}_0 \\ \widehat{\underline{\boldsymbol{c}}}_n - \underline{\boldsymbol{c}}_0 \\ \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}n} - \boldsymbol{\gamma}_{\boldsymbol{\varepsilon}0} \\ \widehat{\boldsymbol{\gamma}}_{\boldsymbol{x}n} - \boldsymbol{\gamma}_{\boldsymbol{x}0} \end{pmatrix} = \boldsymbol{\Psi} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\boldsymbol{\gamma}}_{\boldsymbol{\varepsilon}n} - \boldsymbol{\gamma}_{\boldsymbol{\varepsilon}0} \\ \widehat{\boldsymbol{\gamma}}_{\boldsymbol{x}n} - \boldsymbol{\gamma}_{\boldsymbol{x}0} \end{pmatrix}, \tag{33}$$

where $\mathbf{\Psi}$ is a $d_1 \times d_2$ matrix, $d_1 = 2m^2 + mr + \frac{m(m+1)}{2} + \frac{r(r+1)}{2}$ and $d_2 = md + \frac{m(m+1)}{2} + \frac{r(r+1)}{2}$, given by, for $k = 1, \dots, m$,

$$\mathbf{\Psi}[(k-1)m+1 : km, (k-1)(m+r+2)+2 : (k-1)(m+r+2)+m+1] = I_m,$$
$$\mathbf{\Psi}[m^2+(k-1)m+k, k(m+2)+(k-1)r] = 1,$$
$$\mathbf{\Psi}[2m^2+(k-1)r+1 : m^2+m+kr, k(m+2)+(k-1)r+1 : k(m+r+2)] = I_r,$$
$$\mathbf{\Psi}[2m^2+mr+1 : d_1, md+1 : d_2] = I_{\frac{m(m+1)}{2}+\frac{r(r+1)}{2}}$$

and the others entries are zero. We then have

$$\sqrt{n}\begin{pmatrix} \underline{\widehat{\boldsymbol{\omega}}}_n - \underline{\boldsymbol{\omega}}_0 \\ \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \end{pmatrix} = \widehat{\mathbf{\Omega}}_n \sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \\ \widehat{\gamma}_{\boldsymbol{\varepsilon} n} - \gamma_{\boldsymbol{\varepsilon} 0} \\ \widehat{\gamma}_{\boldsymbol{x} n} - \gamma_{\boldsymbol{x} 0} \end{pmatrix}, \tag{34}$$

where $\widehat{\mathbf{\Omega}}_n = \begin{pmatrix} \widehat{\mathbf{\Omega}}_{1n} \\ \mathbf{\Omega}_2 \end{pmatrix}$ with $\widehat{\mathbf{\Omega}}_{1n} = \begin{pmatrix} \widehat{\boldsymbol{A}}_n^* & \widehat{\boldsymbol{B}}_n^* & \widehat{\boldsymbol{C}}_n^* & \widehat{\boldsymbol{E}}_n^* & \widehat{\boldsymbol{X}}_n^* \end{pmatrix} \mathbf{\Psi}$. Note that $\widehat{\mathbf{\Omega}}_n$ is strongly consistent estimator of $\mathbf{\Omega}$. By applying the Slutzky theorem, the asymptotic normality in Theorem 3 is thus obtained. □

# References


**Bollerslev, T.** (1990) Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *The review of economics and statistics* 495–505.

**Bollerslev, T. and J.M. Wooldridge** (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews* 11, 143–172.

**Boussama, F., F. Fuchs and R. Stelzer** (2011) Stationarity and geometric ergodicity of BEKK multivariate GARCH models.*Stochastic Processes and their Applications* 121, 2331–2360.

**Comte, F. and O. Lieberman** (2003) Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.

**Engle, R.F., V.K. Ng and K.F. Kroner** (1990) Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics* 45, 213–237.

**Engle, R.F. and K.F. Kroner** (1995) Multivariate simultaneous generalized ARCH. *Econometric theory* 11, 122–150.

**Engle, R.F. and Sheppar, K.** (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. *National Bureau of Economic Research.*

**Engle, R.F.** (2002) Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20, 339-350.

**Engle, R.** (2009) *Anticipating correlations: a new paradigm for risk management.* Princeton University Press.

**Francq, C. and J-M. Zakoïan** (1998) Estimating linear representations of nonlinear processes.*Journal of Statistical Planning and Inference* 68, 145–165.

**Francq, C. and J-M. Zakoïan** (2000) Covariance matrix estimation for estimators of mixing weak ARMA models.*Journal of Statistical Planning and Inference* 83, 369–394.

**Francq, C. and J-M. Zakoïan** (2010) *GARCH Models: Structure, Statistical Inference and Financial Applications.* John Wiley.

**Francq, C. and Thieu, L.Q.** (2015) Qml inference for volatility models with covariates. *University Library of Munich, Germany*, No. 63198.

**Francq, C. and G. Sucarrat** (2016) Equation-by-Equation Estimation of a Multivariate Log-GARCH-X Model of Financial Returns. *Journal of the Royal Statistical Society: Series B* 78, 613–635, http://onlinelibrary.wiley.com/doi/10.1111/rssb.12126/epdf

**Hafner, C.M. and A. Preminger** (2009) On asymptotic theory for multivariate GARCH models.*Journal of Multivariate Analysis* 100, 2044–2054.

**Han, H. and D. Kristensen** (2014) Asymptotic Theory for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates. *Journal of Business & Economic Statistics* 32, 416–429.

**Herrndorf, N.** (1984) A functional central limit theorem for weakly dependent sequences of random variables. *The Annals of Probability*, 141–153.

**Pedersen, R.S. and A. Rahbek** (2014) Multivariate variance targeting in the BEKK-GARCH model. *The Econometrics Journal* 17, 24–55.

**Phillips, P.C.B., Y. Sun and S. Jin** (2003) Consistent HAC estimation and robust regression testing using sharp origin kernels with no truncation. *Discussion paper, Yale University.*

**Shephard, N. and Sheppard, K.** (2010) Realizing the future: forecasting with high frequency based volatility (HEAVY) models. *Journal of Applied Econometrics* 25, 197–231.

**Silvennoinen, A. and T. Teräsvirta** (2009) Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. *Journal of Financial Econometrics* 7, 373–411.